DOCUMENT RESUME

ED 053 206 TM 000 738

AUTHOR	Cotton, John W.
TITLE	Mathematical Analysis of a Multiple-Look Concept Identification Model.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	RB-70-72
PUB DATE	Dec 70
NOTE	32p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Mathematical Applications, *Mathematical Models, *Mathematics, *Prediction, *Probability, Probability Theory, Statistical Analysis, *Stimulus Behavior

ABSTRACT

The behavior of focus samples central to the multiple-look model of Trabasso and Bower is examined by three methods. First, exact probabilities of success conditional upon a certain brief history of stimulation are determined. Second, possible states of the organism during the experiment are defined and a transition matrix for those states determined, permitting prediction over all possible numbers of trials. Third, Fisher's generalizations and corrections of the Trabasso and Bower focus sample theory are examined. A general solution for the conditional probability of success is derived from Fisher's equation for the probability of n successes between any two errors. One very strong implication of the theory is given in Section 5. (Author)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

RB-70-72

RESEARCH BULLETIN

ED053206

TM 000 738

MATHEMATICAL ANALYSIS OF A MULTIPLE-LOOK CONCEPT IDENTIFICATION MODEL

John W. Cotton
University of California, Santa Barbara

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Educational Testing Service
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
December 1970

MATHEMATICAL ANALYSIS OF A MULTIPLE-LOOK CONCEPT IDENTIFICATION MODEL

John W. Cotton

University of California, Santa Barbara

Abstract

The behavior of focus samples central to the multiple-look model of Trabasso and Bower (1968) is examined by three methods. First, exact probabilities of success conditional upon a certain brief history of stimulation are determined. Second, possible states of the organism during the experiment are defined and a transition matrix for those states determined, permitting prediction over all possible numbers of trials. Third, Fisher's generalizations and corrections of the Trabasso and Bower focus sample theory are examined. A general solution for the conditional probability of success is derived from Fisher's equation for the probability of n successes between any two errors. One very strong implication of the theory is given in Section 5.

Acknowledgments

The author thanks Dr. John B. Carroll and Dr. Norman Frederiksen for the use of facilities in the Division of Psychological Studies, Educational Testing Service, where this article was prepared. Dr. Carroll, Dr. Roy Freedle, and Mrs. Marilyn Wingersky were generous of their time in discussion of parts of this paper, but bear no responsibility for its final form. Dr. Mary Ann Fisher and Dr. Thomas Trabasso suggested corrections and/or stylistic improvements of the manuscript.

MATHEMATICAL ANALYSIS OF A MULTIPLE-LOOK CONCEPT IDENTIFICATION MODEL

1. Introduction

Consider a K -dimensional binary response concept identification task with one or more dimensions being relevant. A possible solution of the task might be that a stimulus including value 1 of dimension A (A_1) should be followed by Response 1 (R_1) and that A_2 should be followed by R_2. Thus A is a relevant dimension. We require that with more than one relevant dimension all such dimensions give redundant information. Thus, in addition to our assumption about dimension A , we might assume that B_1 must be followed by R_1 , B_2 must be followed by R_2 , and that presentation of A_1 (A_2) always implies presentation of B_1 (B_2) . Thus B is also a relevant dimension. Trabasso and Bower (1968, pp. 54-57) present a model for a focus sample of x relevant and s - x irrelevant cues to which a person may attend on any trial. The focus sample is a crucial part of a multiple look model because it permits the learner to attend to more than one conceivably crucial cue on any one trial, with a subsequent reduction in the number thus noted as new trials give new information.

Trabasso and Bower (1968, p. 54) note that a random sequence of stimulus patterns implies that "each irrelevant cue will have an independent probability p on each trial of being allied with the correct, relevant one." Furthermore, "the probability of the correct response is the proportion of cues in the focus sample that dictate that response." A cue is a dimension value, not a dimension.

Trabasso and Bower begin their derivation by assuming an error on an arbitrarily numbered trial, T_0 . At this point the learner selects s cues,

all of which are consistent with the information from T_0 . Let $\{s\}$ denote the set of such elements, with the convention that we list dimension values which evoke R_1, opposite values evoking R_2 . Thus, if on T_0 $A_1B_1C_1$ should have led to R_1 , $\{A_1,B_1,C_1\}$ is an acceptable set $\{s\}$, for $s = 3$. Trabasso and Bower also permit focus samples in which the same cue appears more than once. For example, $\{s\} = \{B_1,B_1,C_1\}$ is also acceptable in this instance.

On the next trial, T_1 , Trabasso and Bower predict the following proportion of successes:

$$\Pr(S_1|E_0) = \frac{x + (s - x)p}{x + (s - x)} \qquad (1)$$

because x plus $p(s - x)$ is the expected number of cues yielding a correct response, and $x + (s - x) = s$ is the total number of cues from which selection is being made. On subsequent trials in a series of successful trials, any cue which would not have led to a correct response on the immediately previous trial is excluded from the focus sample. The expected number of cues remaining in the focus sample becomes the denominator of a new predictive equation; the expected number of cues which would yield a correct response on the next trial becomes the numerator of that equation. Therefore the following probability is assigned for the $n + 1$ -th success conditional upon n successes in a row following T_0 :

$$\Pr(S_{n+1}|S_n \cdots S_1E_0) = \frac{x + (s - x)p^{n+1}}{x + (s - x)p^n} \quad . \qquad (2)$$

The Trabasso and Bower proof of (2) is brief and appears to be marred by use of the expected operator approximation (Sternberg, 1963, pp. 40-47) without noting

that the expected values just discussed should, after T_1, have been conditionalized subject to successes on all trials up to the point of any prediction. It seems appropriate to make a more rigorous analysis of the consequences of Trabasso and Bower's assumptions. We begin with an examination of specific stimulus sequences.

2. Determination of Response Probabilities for All Stimulus Patterns in a Three-Trial Sequence

The discussion below is dependent upon knowledge of a term from Cotton (in press): Congruence (i) is defined as the number of dimensions, including the relevant one(s) which is (are) consistent with the relevant dimension(s) in changing value(s) from Trial n to Trial $n+1$ when the relevant dimension(s) change(s) or in remaining constant when the relevant dimension(s) remain(s) constant. (The possibility of two or more redundant relevant dimensions is accommodated by the parenthesized s's.)

On tne n-th trial of a K-dimensional binary concept problem there will be 2^{Kn} possible branches reflecting different stimulus sequences of stimuli which may have occurred on the n trials. Though 2^{Kn} is much too many branches to examine explicitly for large n, we can gain useful information by examining a few trials fully in order to determine the possible states of a Markov process and transition probabilities presumed to correspond to the theory in question. Let us consider an example with $K = 3$, one relevant dimension (A_1 should be followed by R_1 and A_2 by R_2), and with $\{s\} = \{A_1,B_1,C_1\}$, one of the acceptable focus samples of size 3 which could follow an error on T_0 for the stimulus $A_1B_1C_1$. We assume for the moment that every one of the eight possible stimulus patterns is equally likely on each

trial, with patterns on pairs of trials having independent probabilities. Under this assumption (and some less restrictive ones), complementary stimuli $A_jB_kC_\ell$ and $A_{j'}B_{k'}C_{\ell'}$, with $j \neq j'$, $k \neq k'$, and $\ell \neq \ell'$ all simultaneously holding as in the case of $A_1B_1C_1$ and $A_2B_2C_2$, have equal probabilities of appearance. Furthermore feedback following presentation of one member of a complementary pair always confirms the same hypotheses which feedback following the other member would confirm. Therefore, in the three-dimensional case it will be sufficient to examine stimulus sequences involving a choice of four stimuli rather than eight. Table 1 shows the possible sequences based on $A_1B_1C_1$, $A_1B_1C_2$, $A_1B_2C_1$, and $A_1B_2C_2$, together with

---

Insert Table 1 about here

---

congruence (i) values, the probability of a correct response (Pr) with each stimulus at each stage, and the {s} values resulting from examining {s} after each success and excluding any dimension value which could have led to an error on that trial. The reader may simply assume that one-half of the events attributed to any stimulus are actually associated with its complementary stimulus.

To read Table 1 easily, one should learn that congruence values (i) for Trial 1 are represented by Roman numerals I, II and III, when cases are delineated on subsequent trials. Case IIA and Case IIB differentiate $i = 2$ cases which involve different stimuli yielding different focus samples. One or two dots following a numerical specification of a case indicates that one or two final trials, respectively, may be ignored as to specific stimulus history because the final focus sample will be independent of that history. Thus for

Case III 1· all focus samples include only the relevant cue following an $i = 3$ trial and an $i = 1$ trial in that order.

Once Table 1 is known, we can use our assumption of equally frequent stimuli to predict $Pr(S_1|E_0)$ with the following equation:

$$Pr(S_1|E_0) = \sum_{seq} wt\ (Pr_1) \qquad (3)$$

where wt is the weight or probability of being in a certain sequence, Pr_n is the probability of a correct response on Trial n (T_n) for that sequence, and the summation is over all sequences. It will be useful to call the right-hand side of (3) by the name $\Sigma\ Prod_1$ and to define:

$$\Sigma\ Prod_{n+1} = \Sigma\ Prod_n\ (Pr_{n+1})$$

$$= \Sigma\ wt\ Pr_1 Pr_2 \cdots Pr_{n+1}$$

$$= Pr(S_{n+1} \cdots S_1|E_0) \quad . \qquad (4)$$

Note that Pr_2 times wt times Pr_1 is the probability of having successes on both T_1 and T_2 during a certain stimulus sequence. It might seem reasonable to let $\Sigma\ Prod_2 = \Sigma\ Pr_2\ wt\ (Pr_1)$ define the probability of two successes in a row after an error without further manipulation. However, the experimental design in question is one in which data on T_2 are not analyzed for subjects making an error on T_1. Therefore, we must take into account the number of subjects remaining for analysis on T_2, i.e., $Pr(S_1|E_0)$ or $\Sigma\ Prod_1$. Since

$$Pr(S_{n+1}|S_n \cdots S_1 E_0) = \frac{Pr(S_{n+1} \cdots S_1|E_0)}{Pr(S_n \cdots S_1|E_0)} \quad , \qquad (5)$$

and

$$\Pr(S_{n+1}|S_n \ldots S_1 E_0) = \frac{\Sigma \text{ Prod}_{n+1}}{\Sigma \text{ Prod}_n} \quad . \tag{6}$$

Table 2 presents the calculations of $\Pr(S_{n+1}|S_n \ldots S_1 E_0)$ for each trial

---

Insert Table 2 about here

---

of the example analyzed in Table 1. Once we determine the value of Trabasso and Bower's p, we can check Table 2 results against (2). First, we emphasize that p is not a response property as in Bower and Trabasso (1964); rather, as the first quotation in this article implies, it is wholly defined once the stimulus probabilities and the reinforcement rule are known. If every irrelevant cue, such as B_1, is exactly as likely to be paired with A_1 as with A_2 in our example, then $p = \frac{1}{2}$. But our assumption of equal probabilities for each possible stimulus pattern assures this equality. Therefore (2) should hold, yielding the same probabilities as obtained in Table 2. It does.

3. A Matrix Formulation of the Focus Sample Problem

Examination of Table 1 suggests that a useful representation of the process under study will result from classification into seven states, with a revised organization leading eventually to four states. The seven states are 1C (the probability of being correct is 1 and all cues in the focus sample are correct); 1U3 (the probability of being correct is 1, but there are three cues in the focus sample, not all of which are correct); 1U2 (like 1U3 but with two cues, not both of which are correct); States 2/3, 1/2, and 1/3 having probabilities of being correct given by their designations; and

the dropout State (D) having zero probability of a correct response because the subject has made an error since T_0 and is therefore excluded from future analyses.

It is easy to identify which state will be operative after a given stimulus sequence by looking at a case number in Table 1 and examining the probability values and {s} entries. Consider Trial 2, Case III: For i values of 3, 2, and 1 a person is in 1U3, 2/3, and 1/3, respectively. Persons making errors on Trial 2 because they are in States 2/3 or 1/3 will go into State D on Trial 3 and stay there thereafter. However, persons who are correct on Trial 2 when in State 2/3 will go into State 1U2 or State 1/2 on Trial 3, depending upon whether the two cues remaining in their focus sample are consistent or inconsistent with the next stimulus presented. Persons who are correct when in State 1/3 on Trial 2 will go into State 1C on Trial 3 since Table 1 shows that only A_1 will remain in their focus sample.

Rather than present a matrix for these seven states, we first expand to 10 states by distinguishing between success (S) and error (E) substates for the three states having fractional probabilities of a correct response. This, together with examinations of probabilities of reaching various points in Table 1, yields the following initial vector:

$$P_1^* = \begin{matrix} 1C & 1U3 & 1U2 & (2/3)S & (2/3)E & (1/2)S & (1/2)E & (1/3)S & (1/3)E & D \\ [\,0 & 1/4 & 0 & 1/3 & 1/6 & 0 & 0 & 1/12 & 1/6 & 0\,] \end{matrix} \qquad (7)$$

and transition matrix:

$$R^* = \begin{array}{c|cccccccccc} & 1C & 1U3 & 1U2 & (2/3)S & (2/3)E & (1/2)S & (1/2)E & (1/3)S & (1/3)E & D \\ \hline 1C & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1U3 & 0 & 1/4 & 0 & 1/3 & 1/6 & 0 & 0 & 1/12 & 1/6 & 0 \\ 1U2 & 0 & 0 & 1/2 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 \\ (2/3)S & 0 & 0 & 1/2 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 \\ (2/3)E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ (1/2)S & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1/2)E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ (1/3)S & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1/3)E & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ D & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \quad . \quad (8)$$

Note that rows 1C, (1/2)S, and (1/3)S of this transition matrix are identical; also rows 1U2 and (2/3)S; also rows (2/3)E, (1/2)E, (1/3)E, and D. By Burke and Rosenblatt's (1958) Corollary 1 we can lump states having such identical rows together, yielding the following 4-state model:

$$P_1 = \begin{array}{c|cccc} & C & U3 & U2 & E \\ \hline & 1/12 & 1/4 & 1/3 & 1/3 \end{array} \quad (9)$$

and

$$R = \begin{array}{c|cccc} & C & U3 & U2 & E \\ \hline C & 1 & 0 & 0 & 0 \\ U3 & 1/12 & 1/4 & 1/3 & 1/3 \\ U2 & 1/4 & 0 & 1/2 & 1/4 \\ E & 0 & 0 & 0 & 1 \end{array} \qquad (10)$$

where C implies that a subject will be correct with probability 1 hereafter; U3 means that a subject will be correct on the current trial but is still unconditioned in that at least one of the cues in the focus sample is irrelevant; U2 means that a subject will be correct on the current trial but that one of the two cues in the focus sample is irrelevant, and E means either that an error will be made on the current trial or that the subject involved has already dropped out of the analysis because of a previous error. The proportion of subjects in the two sources of the E state can be determined by finding the difference between the proportions in E on Trials n and $n - 1$; the difference is the proportion of errors (out of all subjects) on Trial n.

We must now find an expression for R^n in order to obtain explicit trial by trial predictions based on the well known relation:

$$P_{n+1} = P_1 R^n \quad . \qquad (11)$$

A method from Goldberg (1958, pp. 229-231, and exercises 10 and 11, pp. 244-245) leads us first to find the characteristic roots of R from (9) by solving the following determinantal equation:

$$f(\lambda) = |R - \lambda I| = 0 \quad , \qquad (12)$$

where I is an identity matrix, obtaining $\lambda_1 = 1$, $\lambda_2 = 1/4$, $\lambda_3 = 1/2$, and $\lambda_4 = 1$. The Cayley-Hamilton theorem asserts that if $f(\lambda) = 0$ as required by (12), then $f(R)$, using the same constants as in (12) but replacing powers of λ by corresponding powers of R , will equal the null matrix. In the present example, each equation will be a polynomial of the fourth degree.

Now it is possible to write λ^n in the form:

$$\lambda^n = f(\lambda)\, q(\lambda) + r(\lambda) \tag{13}$$

where $q(\lambda)$ is of degree $n - 4$ since $f(\lambda)$ has degree 4 and $r(\lambda)$ has at most degree 3, else $r(\lambda)$ could be factored by $f(\lambda)$. Goldberg cites a proof that the corresponding matrix equation holds as a consequence of (13):

$$R^n = f(R)\, q(R) + r(R) \quad . \tag{14}$$

Invoking the conditions defined by (12) and by the Cayley-Hamilton theorem yields:

$$\lambda^n = r(\lambda) \qquad \text{from (13) and } f(\lambda) = 0$$

$$= a_0 + a_1\lambda + a_2\lambda^2 + a_3\lambda^3 \quad , \tag{15}$$

and

$$R^n = r(R) \quad \text{from (14) and } f(R) = 0$$

$$= a_0 I + a_1 R + a_2 R^2 + a_3 R^3 \tag{16}$$

since $r(\lambda)$ and $r(R)$ must be of degree 3 or less.

We must now solve for the coefficients from (15) and apply them to (16). A slight complication arises because λ_1 and λ_4 are equal, yielding three independent equations, rather than four, from (15). Therefore, we differentiate both sides of (15) with respect to λ, for $\lambda_4 = 1$:

$$n\lambda_4^{n-1} = a_1 + 2a_2\lambda_4 + 3a_3\lambda_4^2 \quad . \tag{17}$$

Substituting the values of λ_1 through λ_4 in (15) or (17), as appropriate, yields the following system of equations:

$$1^n = a_0 + a_1 + a_2 + a_3 \tag{18}$$

$$(1/4)^n = a_0 + a_1/4 + a_2/16 + a_3/64 \tag{19}$$

$$(1/2)^n = a_0 + a_1/2 + a_2/4 + a_3/8 \tag{20}$$

and

$$n = a_1 + 2a_2 + 3a_3 \tag{21}$$

which can also be expressed in matrix form:

$$(c) = C(a) \tag{22}$$

where (c) is the column vector on the left hand side of the set of equations, C is the matrix of coefficients of the a_j's, and (a) is a column vector of a_j's. C is nonsingular; therefore, (22) implies:

$$(a) = C^{-1}(c) \quad . \tag{23}$$

Now inverting C yields:

$$C^{-1} = (1/9)\begin{vmatrix} 13 & 32 & -36 & -3 \\ -88 & -128 & 216 & 21 \\ 164 & 160 & -324 & -42 \\ -80 & -64 & 144 & 24 \end{vmatrix} , \qquad (24)$$

from which (a) has been computed using (23) and has values equal to the coefficients of the R^n terms below based on (16):

$$R^n = (1/9)[13 + 32(1/4)^n - 36(\tfrac{1}{2})^n - 3n]I + (1/9)[-88 - 128(1/4)^n + 216(\tfrac{1}{2})^n + 21n]R$$

$$+ (1/9)[164 + 160(1/4)^n - 324(\tfrac{1}{2})^n - 42n]R^2 + (1/9)[-80 - 64(1/4)^n + 144(\tfrac{1}{2})^n + 24n]R^3 . \qquad (25)$$

We now need values of R^2 and R^3, so that (25) may be applied. By direct calculation, from (10),

$$R^2 = \begin{array}{c|cccc} & C & U3 & U2 & E \\ \hline C & 1 & 0 & 0 & 0 \\ U3 & 9/48 & 1/16 & 1/4 & 1/2 \\ U2 & 3/8 & 0 & 1/4 & 3/8 \\ E & 0 & 0 & 0 & 1 \end{array} \qquad (26)$$

and

$$R^3 = \begin{array}{c|cccc} & C & U3 & U2 & E \\ \hline C & 1 & 0 & 0 & 0 \\ U3 & 49/192 & 1/64 & 7/48 & 7/12 \\ U2 & 7/16 & 0 & 1/8 & 7/16 \\ E & 0 & 0 & 0 & 1 \end{array} . \qquad (27)$$

For present purposes it is sufficient to calculate the last column of R^n, which will be called $(R^n)_4$. Use of (10), (26), and (27) in (25) leads, after simplification, to:

$$(R^n)_4 = \begin{array}{c} \\ C \\ U3 \\ U2 \\ E \end{array} \begin{array}{c} E \\ \left[\begin{array}{c} 0 \\ (2/3)(1 - \frac{1}{2}^n) \\ (\frac{1}{2})(1 - \frac{1}{2}^n) \\ 1 \end{array}\right] \end{array} . \qquad (28)$$

We know from (11) that the probability of being in State E on T_{n+1} is given by:

$$P(E_{n+1}) = P_1(R^n)_4$$

$$= 2/3 - (1/3)\frac{1}{2}^n \quad \text{by (9) and (28).} \qquad (29)$$

But the probability of a success on T_{n+1} is:

$$Pr(S_{n+1}) = 1 - P(E_{n+1})$$

$$= (1/3)(1 + \frac{1}{2}^n) \quad \text{from (29).} \qquad (30)$$

We have just found the probability of a success on T_{n+1}, computed from among all subjects who made an error on T_0. To make this probability conditional upon having been tested on T_{n+1}, we note that we are dealing only with those subjects who were successful on T_1 through T_n and then were also successful on T_{n+1}. Therefore,

$$\Pr(S_{n+1}|S_n \ldots S_1 E_0) = \frac{\Pr(S_{n+1})}{\Pr(S_n)}$$

$$= \frac{1 + \frac{1}{2}^n}{1 + \frac{1}{2}^{n-1}} \quad \text{from (30).} \qquad (31)$$

But (31) is equivalent to (2) for $x = 1$, $s = 3$, and $p = \frac{1}{2}$, the conditions operative in our example. Thus (2) has been verified for the focus sample of Table 1 and equiprobable, independent stimuli.

Extension of the Matrix Formulation to New Focus Samples but the Same Experiment

Trabasso and Bower (1968, pp. 59-60) assume that a subject selects a focus sample by a replacement sampling method in which any one of the K different dimensions has a specific probability of being selected as the first member of the sample, and the same, independent probability of being selected as the second, third, ... or K-th member of the sample. Consequently a focus sample of size s will have from 0 to s elements from any particular dimension. The three-dimensional binary task with $s = 3$ which we have been considering has 10 distinct focus samples, ignoring order, and 27 samples when order is considered. (Other focus samples would be possible if the stimulus on T_0 were different. See Sec. 4.) Table 3 lists the 10 basic focus samples.

Insert Table 3 about here

Sample 10, $\{A_1, B_1, C_1\}$, has already been investigated above. Hopefully a single matrix proof could be developed for (2) which would hold for all 10 samples. Unfortunately Table 3 shows that the rank of the transition matrix

varies from 1 to 4 in the 10 samples under consideration. (In each case the rank is also equal to the dimensionality of the matrix.) Therefore, we have determined an initial vector P_1 and the matrix R for each asterisked sample of Table 3, determined the form of R^n, and verified that in each case (2) follows from use of P_1 and R^n in (11). By symmetry, (2) also holds for each unasterisked sample.

Once (2) or some other equation is known to hold for a focus sample and all possible focus samples have been investigated as above (with the possibility of some samples conforming to different equations or even different forms of equations), the probability of solution of the problem can be determined for each focus sample using (2.2) and the sentence following from Trabasso and Bower (1968, p. 56) and a weighted average probability of solution can be obtained from their (2.3) and (2.4) once one makes a saliency assumption, i.e., specifies the probability of selecting each dimension for use in the focus sample. An equal saliency assumption will, of course, make each of the 27 permutations of Table 3 equally likely.

How Many Trials Must be Examined to Identify the Different States for a Problem with a Specific Focus Sample When s and K Are Large?

The matrix method just presented would be inconvenient if it were necessary to consider all possible stimulus sequences and consequent focus samples in a series longer than the three trials examined above. Suppose K is very large, perhaps 15, and s is even larger, perhaps 20, implying that at least one dimension is represented more than once in the original focus sample. Will this make it necessary to examine more than three trials?

The query just posed may be answered by noting, first, that use of all possible K -dimensional stimuli (excluding complements if desired) on Trial 1 will ensure that all possible combinations of dimensions are retained by various subjects at the end of that trial, excluding possibilities in which the relevant dimension was represented on Trial 1 but not afterward. Thus there will be 1-tuples, 2-tuples, ... K-tuples represented in new focus samples, with the label on a -tuple identifying the number of dimensions represented in a sample, not the number of elements. Because starting with a multiple representation of any dimension can be followed only by keeping all representatives of the dimension or discarding all representatives, no new combinations of dimensions can be produced after Trial 2. But use of all possible stimuli on Trial 2 does enlarge the set of different {s} values by producing all possible consequences on any specific -tuple. Consequently Trial 3 will always include all possible {s} values provided that all possible stimuli were presented on Trial 1 and independently on Trial 2 as well.

4. The Case of Constant Partial Relevance, and Constant Predictability with $Pr \neq .5$

Suppose that, in the example given in Table 1, the four stimuli $A_1B_1C_1$, $A_1B_1C_2$, $A_1B_2C_1$, and $A_1B_2C_2$, were assigned the probabilities .36, .24, .24, and .16 respectively, yielding $Pr(B_1|A_1) = Pr(C_1|A_1) = .6$, so that the partial relevance, p, was constant at .6. [Since the numbering system for B_1, B_2, C_1, and C_2 is arbitrary, reversal of numbers for B_1 and B_2 and for C_1 and C_2 would have yielded $p = 1 - .60 = .40$ for each irrelevant dimension. We adopt the convention of numbering each irrelevant dimension's values so as to maximize each partial relevance, $Pr(B_1|A_1)$, $Pr(C_1|A_1)$,

etc.] Then Table 2 would require a new row of weight values, yielding different values for $Pr(S_1|E_0)$ and related quantities. The new predictions would conform to Eq. 2, showing another case in which Trabasso and Bower's equations hold logically.

Fisher (in press) has shown that in general if the partially relevant hypotheses in the focus sample are divided "into groups according to their probability of producing a correct response (group i will have h_i elements each of which is associated with correct responding with a probability p_i)," then

$$Pr(S_1|E_0) = \frac{x + \Sigma p_i h_i}{x + \Sigma h_i}$$

$$\frac{x + \Sigma p_i h_i}{x + (s - x)} \quad , \tag{32}$$

which reduces to our Eq. 1 if $p = \frac{1}{2}$. Note that two dimensions, B and C, might have the same partial relevance, p, and yet have hypotheses with the same partial relevance ($p_1 = p_2 = p$ for $\{A_1,B_1,C_1\}$ or $p_1 = p_2 = 1 - p$ for $\{A_1,B_2,C_2\}$ or different partial relevances ($p_1 = p$, $p_2 = 1 - p$ for $\{A_1,B_1,C_2\}$). Note also that the i of Eq. 32 is <u>not</u> the congruence value, i, discussed earlier.

If $p_1 = p_2 = p$, Eq. 32 also reduces to Eq. 1, increasing the number of cases in which Trabasso and Bower's conclusions hold. Eq. 2 will also hold in this case, as well as when $p = \frac{1}{2}$.

A case in which Eq. 32 <u>must</u> be employed is easily illustrated by letting the stimulus for T_0 from Table 1 be $A_1B_1C_2$. Since an error was made, one acceptable focus sample is $\{A_1,B_1,C_2\}$. An analogue of Table 2 (not presented) shows that $Pr(S_1|E_0) = .667$, $Pr(S_2|S_1E_0) = .759$, and

$Pr(S_3|S_2S_1E_0) = .842$, for a sequence of trials beginning with this focus sample and $p_1 = .6$, $p_2 = .4$ based on the stimulus probabilities discussed at the beginning of Sec. 4. In contrast Trabasso and Bower [our Eq. (2) with $p = .6$] would have predicted $Pr(S_1|E_0) = .733$, $Pr(S_2|S_1E_0) = .782$, and $Pr(S_3|S_2S_1E_0) = .833)$.

Fisher's Sec. D gives the result:

$$Pr(E_{n+1}S_n \ldots S_1|E_0) = \Sigma p_i^n (1 - p_i) \frac{h_i}{s} \qquad (33)$$

where p_i and h_i are defined as in (32). Eq. (33) can be used to determine

$Pr(S_{n+1}S_n \ldots S_1|E_0)$ or its equivalent from (4), Σ $Prod_{n+1}$:

$$Pr(S_{n+1}S_n \ldots S_1|E_0) + Pr(E_{n+1}S_n \ldots S_1|E_0) = Pr(S_nS_{n-1} \ldots S_1|E_0) \qquad (34)$$

by elementary probability theory. Combining (4) and (34) yields:

$$\Sigma \text{ Prod}_{n+1} = \Sigma \text{ Prod}_n - Pr(E_{n+1}S_n \ldots S_1|E_0) \quad . \qquad (35)$$

Since Σ $Prod_1 = Pr(S_1|E_0)$ from (3) and the discussion following it, (32), (33), and (35) permit a recursion to be performed in order to determine the quantities required to apply (6) for any n .

The method just described may also be applied to the example with a focus sample of $\{A_1, B_1, C_2\}$. The two h_i are each unity, p_1 (for the B variable) is .6, and p_2 (for the C variable) is .4. Equation 33 yields $Pr(E_2S_1|E_0) = .160$ and $Pr(E_3S_2S_1|E_0) = .080$. Equation 32 yields $Pr(S_1|E_0) = .667$; and Eqs. (6) and (35) then imply $Pr(S_2|S_1E_0) = .760$ and $Pr(S_3|S_2S_1E_0) = .842$, these predictions always being within .001 of those reported before for an analogue of Table 2.

5. Further Empirical Implications of the Trabasso-Bower Multiple Look Model

Redundant Relevant Dimensions

Trabasso and Bower developed their model for the specific purpose of treating behavior in the presence of redundant relevant dimensions. The foregoing analyses are in no way changed if we assume k redundant relevant dimensions so that A_1 (A_2) is always also accompanied by A_1' (A_2'),... $A_1^{(k-1)'}$ $[A_2^{(k-1)'}]$. There is no special advantage in discriminating which of the x relevant cues in the initial focus sample comes from each relevant dimension, so we may as well call them all A_1 as in Table 1. Any effect of having relevant redundant x dimensions will be reflected in modifications of the probabilities of the different initial focus samples of size s. Thus for Table 3, equal salience and a single relevant dimension would yield probabilities of 1/3 for each cue to be sampled. Equal salience and k redundant relevant dimensions would yield probabilities of $1/(k + 2)$ for each of the two irrelevant cues and $k/(k + 2)$ for each of the redundant relevant cues to be sampled. Note that each i value in Table 1 is increased by $(k - 1)$ if there are k redundant relevant cues.

Specific Stimulus Sequences

Each of the columns of Table 2 has Pr_1 and Pr_2 values giving the probability of a success on T_1 and a subsequent success on T_2 for specific stimulus values presented in sequence, as well as Pr_3 values giving the average probability of success on T_3 following the sequence of T_1 and T_2, conditional on success on both previous trials. Tables 1 and 2 could be expanded for larger n in order to treat longer stimulus sequences. However, a more convenient method is to find a sequence of matrices comparable to that

of (10), with each one appropriate to the stimulus on a certain trial, applying them in series:

$$P_{n+1} = P_1 R_1 R_{T_2} \cdots R_{T_n} \tag{36}$$

where R_{T_j} is the transition matrix appropriate to the stimulus change from T_j to T_{j+1}. This method of prediction is illustrated in detail in Cotton (in press), using a single-look model.

A very severe test of the present model is suggested by examination of Table 1 for congruence values (i) of 1: First, consider the case in which $x \geq 1$. Among all subjects who erred on T_0 and had $i = 1$ in Table 1 (or had i = the number of relevant dimensions for a more general case) on T_1 and who were successful on T_1, none will keep an irrelevant cue on T_2 because no irrelevant cue placed in the sample focus on T_0 can be consistent with the relevant cue(s) on T_1, by the definition of congruence. Thus none of the subjects with this history will ever again make an error on this problem.

Now consider the case in which $x = 0$. For example, let the stimulus on T_0 be $A_1B_1C_1$ and $\{s\} = \{B_1, B_1, C_1\}$ be the focus sample selected to be consistent with reinforcement of R_1 on T_0, with A being the relevant dimension. On T_1, for which $i = 1$, the B and C dimension values on T_1 will both be inconsistent with the value of the A dimension on T_1. Therefore, the probability of a correct answer on T_1 will be zero. This conclusion holds for any case in which $x = 0$. Consequently, <u>all subjects who err on T_0, have i = the number of relevant dimensions on T_1, and are successful on T_1, will be errorless ever after, according to the multiple-look model</u>. This implication can be expanded to permit the $i = 1$ successful trial to occur after T_1; we do not examine the logic of that case here. Failure of this prediction

is equivalent to failure of a strict "local consistency" theory (Gregg & Simon, 1967).

We do not know of a published set of data bearing upon this prediction. However, Pyle (1969) performed two experiments in which his Group 1 had $i = 1$ on Trial 2 and on all subsequent trials except those numbered with multiples of 5. Raw data kindly provided by Pyle show that in Experiment 1 only 12 of 18 subjects with a success on an $i = 1$ trial following an error made no further errors. The corresponding result for Experiment 2 was 20 out of 31.

Cotton and Rhone (1970) have performed an experiment in which Group 1 has the same i values as in Pyle's two Group 1's. Among 23 subjects in Cotton and Rhone's Group 1, 18 had an error on T_0 and a success on T_1, for some arbitrary T_1 not equal to a multiple of 5. Of these 18, 9 made no further errors in the 24-trial sequence given all subjects. Thus 9 out of 18 subjects exhibited behavior flatly contradicting the strong prediction just derived from the multiple-look model.

It is easy to show that the prediction of errorless performance once a correct response is given with i = number of relevant dimensions (assuming an error on the previous trial) also follows from Trabasso and Bower's (1968, pp. 219-226) modified multiple-look model. That model assumes that, following a correct response, the subject has probability b of excluding inappropriate hypotheses on the same basis as the original model and probability $1 - b$ of excluding them but resampling from locally consistent hypotheses in order to keep the size of s constant. The hypothesis which has been in the focus sample for the greatest number of correct trials is called the dominant hypothesis and will control the response on any given trial. Since the i = number

of relevant dimensions condition assures that the correct hypothesis or hypotheses will be the only one(s) in $\{s\}$ on T_1 which were also confirmed on the error trial T_0, previous to resampling, the correct hypothesis or hypotheses will be the only one(s) in $\{s\}$ on T_1 which were also confirmed on the error trial T_0, previous to resampling, the correct hypothesis or hypotheses will be the dominant one(s) on the next trial, will again be confirmed and still be dominant, etc., assuring no subsequent errors.

Prediction of the Distribution of Runs of All Successes or All Errors

Trabasso and Bower (1968, pp. 55-56) derive an equation for the probability of a run of n successive successes following an error: $Pr(H = n) = (1 - p)p^n$. For all focus samples for which Eq. 2 holds, Trabasso and Bower's formula for $Pr(H = n)$ stands as given. Since this formula does not depend directly upon either x or s, a subject could shift from one acceptable focus sample to another following each error (as he is assumed to do by the theory) and yet the same equation would hold throughout his session, permitting calculation of a variety of run statistics such as those presented in Bower and Trabasso (1964) for a single-look model. We emphasize a point inherent in Trabasso and Bower's discussion: The case $s - x = 0$ is acceptable for a focus sample because it will produce learning, making $Pr(H = \infty) = 1$ at the end of the experiment. However, this serves to emphasize that the learning parameter, x/s, defined in their (2.2), is most assuredly not constant within a session for a single s but rather ranges from 0 when $x = 0$ to 1 when $x = s$.

For the general case, Fisher (in press) has shown that $Pr(H = n) = \frac{1}{\Sigma h_i}[\Sigma p_i^n(1 - p_i)h_i]$, using the same notation as in Eq. 32. This equation reduces to the Trabasso and Bower result for any case in which Eqs. 1 and 2

hold. A final thought about this general case: One error may occur in response to $A_1B_1C_1$, as in Table 1; the next error may occur in response to $A_1B_1C_2$ as in our later example, so that $Pr(H = n)$ must be computed separately for each case because the p_i values will not be constant throughout the experiment even though the partial relevance of any cue is constant. Introduction of sampling schemes for focus samples, as in Trabasso and Bower (1968, pp. 57-60) must receive careful mathematical analysis since this problem of shifting p_i values has not previously been noted.

6. Summary

Two methods of deriving predictions for the Trabasso and Bower multiple-look concept identification model have been examined. A method of directly calculating the effects of every possible stimulus sequence is practical only for small numbers of trials and must be used separately for each possible focus sample of a given size. However, it can be employed for cases of partially relevant cues, redundant relevant cues, and a single stimulus sequence for all subjects. This method reveals a very strong implication of the model: Among subjects who make an error on some trial T_0 and who are correct on the immediately subsequent trial for which the congruence must be equal to the number of relevant dimensions, there will be no further errors. Existing data on this point contradicts the theory.

A matrix method of proof is applicable for all trial numbers and is otherwise comparable to the first method. Use of the first method for three trials will normally be necessary to determine the appropriate transition matrix, which varies from one focus sample to another.

This paper also discusses Fisher's demonstration that certain Trabasso and Bower equations sometimes fail to hold if $p \neq .50$. Her conclusions are shown to imply a general procedure for calculating the probability of a success on Trial n given success on all previous trials since an error on Trial 0.

References

Bower, G. H., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford: Stanford University Press, 1964.

Burke, C. J., & Rosenblatt, M. A Markovian function of a Markov chain. Annals of Mathematical Statistics, 1958, 29, 1112-1122.

Cotton, J. W. A sequence-specific concept identification model: Infrastructure for the Bower and Trabasso theory. Journal of Mathematical Psychology, in press.

Cotton, J. W., & Rhone, S. Concept identification as a function of congruence value. Paper read at Mathematical Psychology Meetings, Miami Beach, Florida, Sept. 1970.

Fisher, M. A. A note on the generalization of some results to hypothesis sampling theory. Journal of Mathematical Psychology, in press.

Goldberg, S. Introduction to difference equations. New York: Wiley, 1958.

Gregg, L. W., & Simon, H. A. Process models and stochastic theories of simple concept formation. Journal of Mathematical Psychology, 1967, 4, 246-276.

Pyle, T. W. Sequential aspects of concept identification. Unpublished Ph.D. dissertation, University of Texas, 1969.

Sternberg, S. Stochastic learning theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology, Vol. 2. New York and London: Wiley, 1963.

Trabasso, T., & Bower, G. H. Attention in learning: Theory and research. New York, London, and Sydney: Wiley, 1968.

Table 1

An Examination of Stimulus Patterns, Congruence Values (i), Success Probabilities (Pr)

Conditional on Stimulus Patterns, and Attendant Focus Samples {s}, with s = 3

[Solution: $A_1 \rightarrow R_1$, $A_2 \rightarrow R_2$. On T_0, $S = A_1B_1C_1$. R_2 occurred and was incorrect. {s} = $\{A_1,B_1,C_1\}$ on T_1 is an acceptable consequence of the T_0 events and will be studied here.]

	Trial 1			Trial 2								
				Case IIA			Case IIB			Case III		
Stim	i	Pr	{s} after correct	i	Pr	{s} after correct	i	Pr	{s} after correct	i	Pr	{s} after correct
$A_1B_1C_1$	3	1	A_1,B_1,C_1	2a	1	A_1,B_1	2a	1	A_1,C_1	3	1	A_1,B_1,C_1
$A_1B_1C_2$	2A	2/3	A_1,B_1	3	1	A_1,B_1	1	$\frac{1}{2}$	A_1	2a	2/3	A_1,B_1
$A_1B_2C_1$	2B	2/3	A_1,C_1	1	$\frac{1}{2}$	A_1	3	1	A_1,C_1	2b	2/3	A_1,C_1
$A_1B_2C_2$	1	1/3	A_1	2b	$\frac{1}{2}$	A_1	2b	$\frac{1}{2}$	A_1	1	1/3	A_1

	Trials 2 and 3			Trial 3								
	Case I··			Case IIA1·			Case IIB1·			Case III1·		
Stim	i	Pr	{s} after correct	i	Pr	{s} after correct	i	Pr	{s} after correct	i	Pr	{s} after correct
All	All	1	A_1	All	1	A_1	All	1	A_1	All	1	A_1

Table 1 (continued)

Stim	Case IIA2a i	Pr	{s} after correct	Case IIB2a i	Pr	{s} after correct	Case III2a i	Pr	{s} after correct
$A_1B_1C_1$	3	1	A_1,B_1	3	1	A_1,C_1	2a	1	A_1,B_1
$A_1B_1C_2$	2a	1	A_1,B_1	2a	$\frac{1}{2}$	A_1	3	1	A_1,B_1
$A_1B_2C_1$	2b	$\frac{1}{2}$	A_1	2b	1	A_1,C_1	1	$\frac{1}{2}$	A_1
$A_1B_2C_2$	1	$\frac{1}{2}$	A_1	1	$\frac{1}{2}$	A_1	2b	$\frac{1}{2}$	A_1

Stim	Case IIA2b i	Pr	{s} after correct	Case IIB2b i	Pr	{s} after correct	Case III2b i	Pr	{s} after correct
$A_1B_1C_1$	1	1	A_1	1	1	A_1	2a	1	A_1,C_1
$A_1B_1C_2$	2a	1	A_1	2a	1	A_1	1	$\frac{1}{2}$	A_1
$A_1B_2C_1$	2b	1	A_1	2b	1	A_1	3	1	A_1,C_1
$A_1B_2C_2$	3	1	A_1	3	1	A_1	2b	$\frac{1}{2}$	A_1

Stim	Case IIA3 i	Pr	{s} after correct	Case IIB3 i	Pr	{s} after correct	Case III3 i	Pr	{s} after correct
$A_1B_1C_1$	2a	1	A_1,B_1	2a	1	A_1,C_1	3	1	A_1,B_1,C_1
$A_1B_1C_2$	3	1	A_1,B_1	1	$\frac{1}{2}$	A_1	2a	2/3	A_1,B_1
$A_1B_2C_1$	1	$\frac{1}{2}$	A_1	3	1	A_1,C_1	2b	2/3	A_1,C_1
$A_1B_2C_2$	2b	$\frac{1}{2}$	A_1	2b	$\frac{1}{2}$	A_1	1	1/3	A_1

Table 2

Computation of $Pr(S_1|E_0)$, $Pr(S_2|S_1E_0)$, and $Pr(S_3|S_2S_1E_0)$ from Equations (3) through (5) and the Data of Table 1

(Some averaging across cases has been performed to reduce the size of this table. Weights follow from equal probabilities for each stimulus and independence of stimuli on different trials.)

Case	I··	IIA1·	IIA2a·	IIA2b·	IIA3·	IIB1·	IIB2a·	IIB2b·	IIB3·	III1·	III2a·	III2b·	III3·	Sum
Weight	1/4	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1
Pr_1	1/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	2/3	1	1	1	1	-
$Product_1$	1/12	1/24	1/24	1/24	1/24	1/24	1/24	1/24	1/24	1/16	1/16	1/16	1/16	2/3
Pr_2	1	1/2	1	1/2	1	1/2	1	1/2	1	1/3	2/3	2/3	1	-
$Product_2$	1/12	1/48	1/24	1/48	1/24	1/48	1/24	1/48	1/24	1/48	1/24	1/24	1/16	1/2
Pr_3	1	1	3/4	1	3/4	1	3/4	1	3/4	1	3/4	3/4	2/3	-
$Product_3$	1/12	1/48	1/32	1/48	1/32	1/48	1/32	1/48	1/32	1/48	1/32	1/32	1/24	5/12

$$Pr(S_1|E_0) = \text{Sum Product}_1 = 2/3$$

$$Pr(S_2|S_1E_0) = \frac{\text{Sum Product}_2}{\text{Sum Product}_1} = \frac{1/2}{2/3} = 3/4$$

$$Pr(S_3|S_2S_1E_0) = \frac{\text{Sum Product}_3}{\text{Sum Product}_2} = \frac{5/12}{1/2} = 5/6$$

Table 3

Listing of All Possible Initial Focus Samples in a Three-Dimensional Problem with $s = 3$, $S = A_1B_1C_1$ on T_0, and Dimension "A" Relevant

Focus Sample	{s}	x	No. Permutations	Rank of R Matrix
1*	$\{A_1,A_1,A_1\}$	3	1	1
2*	$\{B_1,B_1,B_1\}$	0	1	2
3	$\{C_1,C_1,C_1\}$	0	1	2
4*	$\{B_1,B_1,C_1\}$	0	3	4
5	$\{B_1,C_1,C_1\}$	0	3	4
6*	$\{A_1,A_1,B_1\}$	2	3	3
7	$\{A_1,A_1,C_1\}$	2	3	3
8*	$\{A_1,B_1,B_1\}$	1	3	3
9	$\{A_1,C_1,C_1\}$	1	3	3
10*	$\{A_1,B_1,C_1\}$	1	6	4
			Sum = 27	

*By symmetry, any unasterisked sample behaves like the asterisked sample above it. Only asterisked samples were explicitly investigated.